

16

Using IBM SPSS for Windows

Chapter guide

Introduction	354
Getting started in SPSS	355
Beginning SPSS	355
Entering data in the Data Viewer	356
Defining variables: variable names, missing values, variable labels, and value labels	357
Recoding variables	359
Computing a new variable	359
Data analysis with SPSS	361
Generating a frequency table	361
Generating a bar chart	363
Generating a pie chart	363
Generating a histogram	363
Generating the arithmetic mean, median, standard deviation, the range, and boxplots	363
Generating a contingency table, chi-square, and Cramér's V	366
Generating Pearson's r and Spearman's ρ	368
Generating scatter diagrams	368
Comparing means and eta	372
Generating a contingency table with three variables	372
Further operations in SPSS	373
Saving your data	373
Retrieving your data	374
Printing output	374
Key points	374
Questions for review	374



Chapter guide

In order to implement the techniques that you learned in Chapter 15, you would need to do either of two things: learn the underlying formula for each technique and apply your data to it, or use computer software to analyse your data. The latter is the approach chosen in this book for two main reasons.

- It is closer to the way in which quantitative data analysis is carried out in real research nowadays.
- It helps to equip you with a useful transferable skill.

You will be learning SPSS for Windows, which is the most widely used package of computer software for doing this kind of analysis. It is relatively straightforward to use. I will be continuing to refer to the techniques introduced in Chapter 15 and will continue to use the gym survey as an example.

This chapter largely operates in parallel to Chapter 15, so that you can see the links between the techniques learned there and the use of SPSS to implement them.

Introduction

This chapter aims to provide a familiarity with some basic aspects of SPSS for Windows, which is possibly the most widely used computer software for the analysis of quantitative data for social scientists. SPSS, which originally was short for Statistical Package for the Social Sciences, has been in existence since the mid-1960s and over the years has undergone many revisions, particularly since the arrival of personal computers. The version that was used in preparing this chapter was Release 18 and is called PASW Statistics 18. At the time of writing this chapter Release 19 was being beta tested but the indications were that the differences between Releases 18 and 19 were, from the point of view of the material covered in this chapter, largely to do with the appearance of some of the dialog boxes and menus that appear. The beta version is referred to as IBM SPSS Statistics Version 19. From this point on, when referring to SPSS for Windows or PASW Statistics 18 in the text, it will be

called simply SPSS and Release 18 will be the focus of attention. The gym survey used in Chapter 15 will be employed to illustrate SPSS operations and methods of analysis. The aim of this chapter is to introduce ways of using SPSS to implement the methods of analysis discussed in Chapter 15.

SPSS operations will be presented in **bold**, for example, **Variable Name**: and **Analyze**. Names given to variables in the course of using SPSS will be presented in **bold italics**, e.g. *gender* and *reasons*. Labels given to values or to variables are also in bold but in a different font, e.g. **reasons for visiting** and **male**. Tips and skills 'Basic operations in SPSS' presents a list summarizing these. One further element in the presentation is that a right-pointing arrow (→) will be used to denote 'click once with the left-hand button of your mouse'. This action is employed to make selections and similar activities.



Tips and skills

Basic operations in SPSS

- The **SPSS Data Editor**. This is the sphere of SPSS into which data are entered and subsequently edited and defined. It is made up of two screens: the **Data Viewer** and the **Variable Viewer**. You move between these two viewers by selecting the appropriate tab at the bottom of the screen.
- The **Data Viewer**. This is the spreadsheet into which your data are entered. When you start up SPSS, the **Data Viewer** will be facing you.

- The **Variable Viewer**. This is another spreadsheet, but this one displays information about each of the variables and allows you to change that information. It is the platform from which you provide for each variable such information as: the variable name; a variable label; and value labels (see below).
- The **Output Viewer**. When you perform an analysis or produce a diagram (called a 'chart' in SPSS), your output will be deposited here. The **Output Viewer** superimposes itself over the **Data Editor** after an analysis has been performed or a chart generated.
- A **Variable Name**. This is the name that you give to a variable, e.g. *gender*. The name must be no more than eight characters. Until you give a variable a name, it will be referred to as *var00001*, etc. When the variable has been given a name, it will appear in the column for that variable in the Data View window. It is generated from the Variable Viewer.
- A **Variable Label**. This is a label that you can give to a variable but which is not restricted to eight characters. Spaces can be used, e.g. **reasons for visiting**. The Label will appear in any output you generate. It is generated from the **Variable Viewer**.
- A **Value Label**. This is a label that you can attach to a code that has been used when entering data for all types of variables other than interval/ratio variables. Thus, for *var00001*, we would attach the label male to 1 and female to 2. When you generate output, such as a frequency table or chart, the labels for each value will be presented. This makes the interpretation of output easier. It is generated from the Variable Viewer.
- **Missing Values . . .** When you do not have data for a particular variable when entering data for a case, you must specify how you are denoting missing values for that variable. Missing values are generated from the Variable Viewer.
- **Recode**. A procedure that allows codes or numbers to be changed. It is especially helpful when you need to combine groups of people—for example, when producing age bands.
- **Compute . . .** A procedure that allows you to combine two or more variables to form a new variable.
- **Analyze**. This is the point on the menu bar above the **Data Editor** from which you choose (via a dropdown menu) which method of analysis you want to select. Note that, whenever an item on a menu appears with a right-pointing arrowhead after it, this means that, if you select that option, a further menu will follow on.
- **Graphs**. This is the point on the menu bar above the **Data Editor** from which you choose (via a drop-down menu) which chart you want to select.
- **Chart Editor**. When you produce a graph, you can edit it with the **Chart Editor**. To activate this editor, double-click anywhere in the graph. A small chart editor window will appear and your main graph will appear opaque until you exit the Editor. From the Editor, you can make various changes and enhancements to your graph.



Getting started in SPSS

Beginning SPSS

To start SPSS, double-click on the **PASW Statistics** icon on your computer screen. If there is no icon, → the Start button in the bottom left-hand corner of your screen. From the menu of programs, → **SPSS Inc.** A follow-on menu will appear, from which you should select **PASW Statistics 18**. When SPSS loads, you may be faced with an opening dialog box with the title 'What do you want to do?' and a list of options. Many users prefer to disable

this opening box. It is not important in relation to the following exposition, so → **Cancel**. You will then be faced with the **SPSS Data Editor**. This is made up of two components: **Data View** and **Variable View**. In the following discussion, these two screens are referred to as the **Data Viewer** and the **Variable Viewer**. You move between these two viewers by selecting the appropriate tab at the bottom of the screen. The **Data Viewer** is in the form of a spreadsheet grid into which you enter your data. The columns represent *variables* – in other words,

information about characteristics of each person in the gym study sample. Until data are entered and names are given to variables, each column simply has **var** as its heading. The rows represent **cases**, which can be people (as in the example you will be working through) or any unit of analysis. Each block in the grid is referred to as a 'cell'. Note also that when the data are in the SPSS spreadsheet, they will look different; for example, 1 will be 1.00.

Entering data in the Data Viewer

To input the data into the **Data Viewer**, make sure that the top left-hand cell in the grid is highlighted (see Plate 16.1). If it is not highlighted, simply click once in that cell. Then, type the appropriate figure for that cell—that is, 1. This number goes directly into that cell and into the box beneath the toolbar. As an alternative to using the mouse, many people find it easier to use the arrow keys on their keyboard to move from cell to cell. If you make a mistake at any point, simply click once in the cell in question, type in the correct value, and click once more

in that cell. When you have finished, you should end up in the bottom right-hand cell of what will be a perfect rectangle of data. Plate 16.2 shows the **Data Viewer** with the data from the gym survey entered (though only part of the set of data is visible, in that only the first thirty-one respondents are visible). The first row of data contains the coded answers from the completed questionnaire in Chapter 15 (see Tips and skills 'A completed and processed questionnaire').

In order to proceed further, you will find that SPSS works in the following typical sequence for defining variables and analysing your data.

1. You make a selection from the menu bar at the top of the screen, e.g. → **Analyze**.
2. From the menu that will appear, make a selection, e.g. → **Descriptive Statistics**.
3. This will bring up a **dialog box** into which you will usually inform SPSS of what you are trying to do—e.g. which variables are to be analysed.
4. Very often, you then need to convey further information and to do this you have to → a button that will

Plate 16.1

The SPSS Data Viewer

Each row represents a case

Each column represents a variable

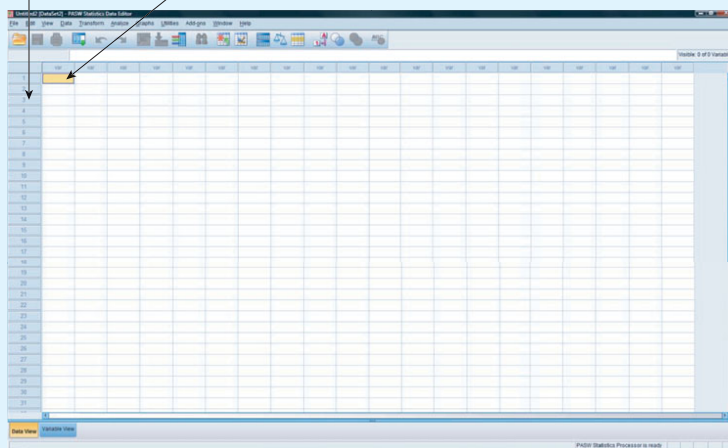


Plate 16.2

The Data Viewer with 'gym study' data entered

This row shows the data for the first person who answered the Gym Survey questionnaire

	var00001	var00002	var00003	var00004	var00005	var00006	var00007	var00008	var00009	var00010	var00011	var00012
1	1.00	27.00	2.00	1.00	3.00	1.00	2.00	0	33.00	17.00	0.00	0.00
2	2.00	44.00	1.00	3.00	1.00	4.00	3.00	1.00	2.00	10.00	23.00	10.00
3	2.00	19.00	3.00	1.00	2.00	2.00	1.00	1.00	1.00	27.00	18.00	12.00
4	2.00	27.00	3.00	2.00	1.00	2.00	1.00	2.00	0	38.00	10.00	3.00
5	1.00	17.00	2.00	1.00	3.00	2.00	3.00	1.00	4.00	22.00	0	10.00
6	2.00	27.00	3.00	1.00	1.00	3.00	1.00	3.00	3.00	34.00	17.00	0
7	1.00	39.00	1.00	2.00	1.00	5.00	1.00	1.00	5.00	17.00	48.00	10.00
8	2.00	36.00	3.00	1.00	2.00	2.00	2.00	1.00	1.00	25.00	18.00	7.00
9	1.00	37.00	2.00	1.00	1.00	3.00	1.00	2.00	0	34.00	18.00	0
10	2.00	51.00	2.00	2.00	2.00	4.00	3.00	2.00	0	16.00	18.00	15.00
11	1.00	24.00	1.00	2.00	1.00	3.00	1.00	1.00	1.00	0	42.00	16.00
12	2.00	29.00	2.00	1.00	2.00	3.00	1.00	2.00	0	34.00	22.00	12.00
13	1.00	39.00	1.00	1.00	1.00	2.00	1.00	2.00	0	23.00	24.00	7.00
14	2.00	22.00	2.00	1.00	3.00	4.00	2.00	1.00	3.00	37.00	14.00	12.00
15	0.00	44.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16	2.00	41.00	3.00	1.00	2.00	2.00	3.00	1.00	4.00	22.00	7.00	10.00
17	1.00	25.00	1.00	1.00	1.00	3.00	1.00	1.00	1.00	21.00	29.00	4.00
18	2.00	48.00	3.00	1.00	2.00	4.00	2.00	1.00	4.00	19.00	8.00	11.00
19	1.00	30.00	3.00	1.00	1.00	5.00	1.00	2.00	0	23.00	9.00	6.00
20	1.00	25.00	1.00	2.00	1.00	3.00	1.00	1.00	1.00	23.00	19.00	0
21	2.00	24.00	2.00	1.00	1.00	3.00	2.00	1.00	2.00	29.00	7.00	6.00
22	2.00	39.00	1.00	2.00	3.00	5.00	1.00	0	17.00	0	9.00	0
23	1.00	44.00	3.00	1.00	1.00	3.00	2.00	1.00	2.00	22.00	8.00	5.00
24	1.00	0	1.00	2.00	2.00	4.00	2.00	1.00	4.00	15.00	10.00	4.00
25	2.00	18.00	3.00	1.00	2.00	3.00	1.00	2.00	1.00	16.00	7.00	10.00
26	1.00	41.00	3.00	1.00	1.00	3.00	1.00	2.00	0	34.00	10.00	4.00
27	2.00	38.00	2.00	1.00	2.00	5.00	3.00	1.00	2.00	24.00	14.00	10.00
28	1.00	26.00	2.00	1.00	2.00	1.00	1.00	2.00	0	48.00	22.00	7.00
29	1.00	41.00	1.00	2.00	1.00	3.00	1.00	1.00	2.00	17.00	27.00	0
30	2.00	38.00	3.00	1.00	1.00	2.00	2.00	2.00	0	32.00	13.00	10.00
31	2.00	29.00	3.00	1.00	3.00	2.00	1.00	2.00	0	31.00	0	7.00

bring up what is called, following Bryman and Cramer (2011), a *sub-dialog box*.

5. You then provide the information in the sub-dialog box and then go back to the dialog box. Sometimes, you will need to bring up a further sub-dialog box and then go back to the dialog box.

When you have finished going through the entire procedure, → **OK**. The toolbar beneath the menu bar allows shortcut access to certain SPSS operations.

Defining variables: variable names, missing values, variable labels, and value labels

Once you have finished entering your data, you need to define your variables. The following steps will allow you to do this:

1. → the **Variable View** tab at the bottom of the **Data Viewer** (opens the **Variable Viewer** shown in Plate 16.3).

2. To provide a variable name, click on the current variable name (e.g. **var00003**) and type the name you want to give it (e.g. **reasons**). Remember that this name must be no more than eight characters and you *cannot* use spaces.
3. You can then give your variable a more detailed name, known in SPSS as a variable label. To do this, → cell in the **Label** column relating to the variable for which you want to supply a variable label. Then, simply type in the variable label (i.e. **reasons for visiting**).
4. Then you will need to provide 'value labels' for variables that have been given codes. The procedure generally applies to variables that are not interval/ratio variables. The latter, which are numeric variables, do not need to be coded (unless you are grouping them in some way). To assign value labels, → in the **Values** column relating to the variable you are working on. A small button with three dots on it will appear. → the button. The **Value Labels** dialog box will appear (see Plate 16.4). → the box to the right of **Value** and begin to define the value labels. To do this, enter the value

Plate 16.3

The Variable Viewer

To create **Missing Values** for **var00003**, click here. A little button with 3 dots will appear

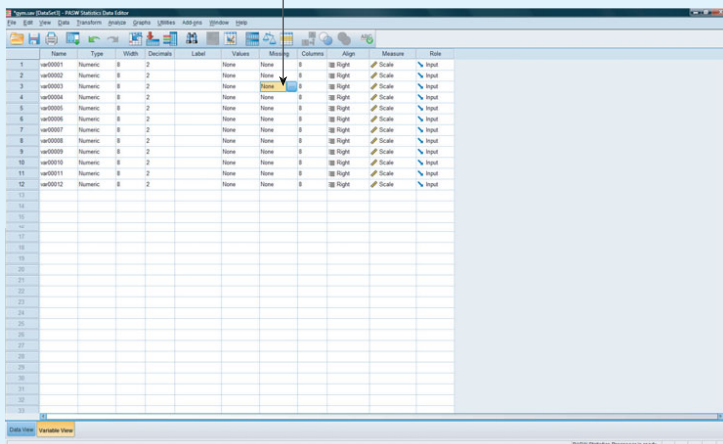


Plate 16.4

The Value Labels dialog box

Remember to click on **Add** after entering each **Value** and **Label**

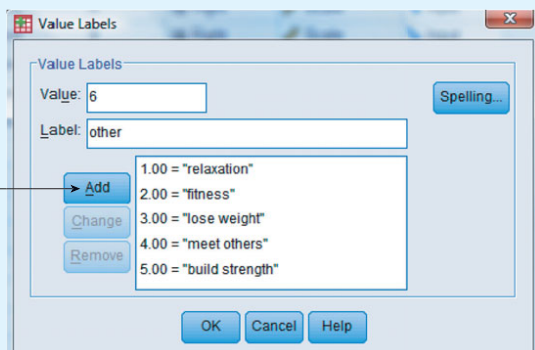
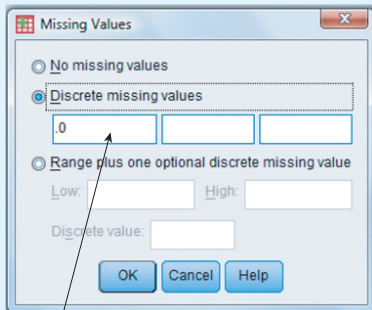


Plate 16.5

The Missing Values dialog box



Designates 0 as the missing value for the variable in question

(e.g. 1) in the area to the right of **Value** and then the value label (e.g. *relaxation*) in the area to the right of **Label**. Then → **Add**. Do this for each value. When you have finished → **OK**.

5. You will then need to inform SPSS of the value that you have nominated for each variable to indicate a missing value. In the case of *reasons*, the value is 0 (zero). To assign the missing value, → the cell for this variable in the **Missing** column. Again, → the button that will appear with three dots on it. This will generate the **Missing Values** dialog box (see Plate 16.5). In the **Missing Values** dialog box, enter the missing value (0) below **Discrete missing values**: and then → **OK**.

In order to simplify the following presentation, *reasons* will be the only variable for which a variable label will be defined.

Recoding variables

Sometimes you need to recode variables—for example, when you want to group people. You would need to do this in order to produce a table like Table 15.3 for an interval/ratio variable like **var00002**, which we will give the variable name *age*. SPSS offers two choices: you can recode *age* so that it will be changed in the **Data Viewer**, or you can keep *age* as it is and create a new variable. This latter option is desirable whenever you want to pre-

serve the variable in question as well as create a new one. Since we may want to carry out analyses involving *age* as an interval/ratio variable, we will recode it so that a new variable, which we will call *agegp*, for *age groups*, will be created. The aim of the following operations is to create a new variable—*agegp*—which will comprise five age bands, as in Table 15.3.

1. → **Transform** → **Recode** → **Into Different Variables ...** [opens **Recode into Different Variables** dialog box shown in Plate 16.6]
2. → *age* → [puts *age* in **Numeric Variable -> Output Variable Name**: and type *agegp* → **Change** [puts *agegp* in the **Numeric Variable -> Output Variable**: box] → **Old and New Values ...** [opens **Recode into Different Variables: Old and New Values** sub-dialog box shown in Plate 16.7]
3. → the circle by **System- or user-missing** and by **System-missing** under **New Value**, if you have missing values for a variable, which is the case for this variable
4. → circle by **Range, LOWEST through value**: and type **20** in the box → box by **Value** under **New Value** and type **1** → **Add** [the new value will appear in the **Old -> New**: box]
5. → first box by **Range**: and type **21** and in box after **through** type **30** → box by **Value** under **New Value** and type **2** → **Add**
6. → first box by **Range**: and type **31** and in box after **through** type **40** → box by **Value** under **New Value** and type **3** → **Add**
7. → first box by **Range**: and type **41** and in box after **through** type **50** → box by **Value** under **New Value** and type **4** → **Add**
8. → circle by **Range, value through HIGHEST** and type **51** in the box → box by **Value** in **New Value** and type **5** → **Add** → **Continue** [closes the **Recode into Different Variables: Old and New Values** sub-dialog box shown in Plate 16.7 and returns you to the **Recode into Different Variables** dialog box shown in Plate 16.6]
9. → **OK**

The new variable *agegp* will be created and will appear in the **Data Viewer**. You would then need to generate **value labels** for the five age bands and possibly a **variable label** using the approach described above.

Computing a new variable

A person's total amount of time spent in the gym is made up of three variables: *cardmins*, *weimins*, and *othmins*.

Plate 16.6

The Recode into Different Variables dialog box

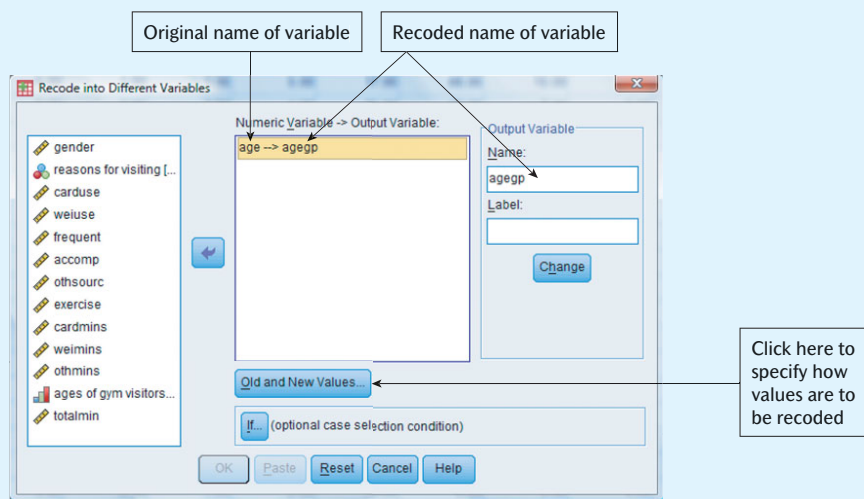


Plate 16.7

The Recode into Different Variables: Old and New Values sub-dialog box

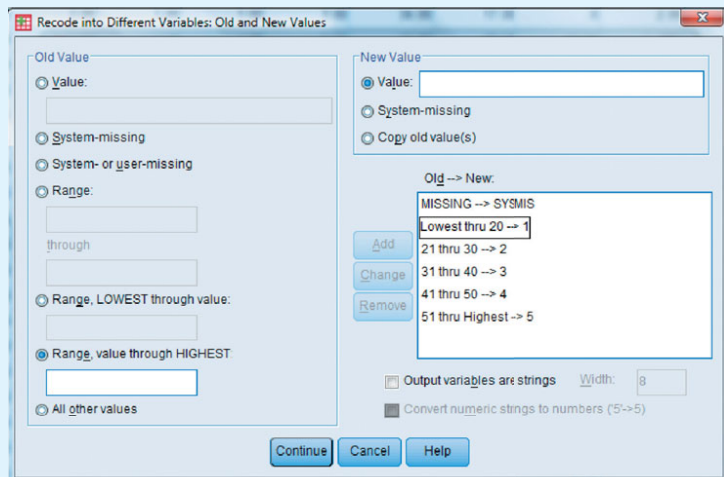
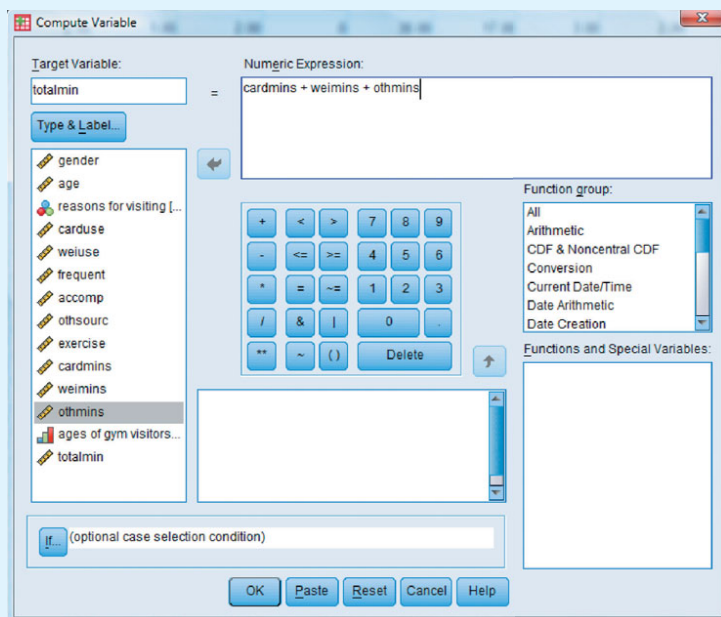


Plate 16.8

The Compute Variable dialog box



If we add these up, we should arrive at the total number of minutes spent on activities in the gym. In so doing, we will create a new variable **totalmin**. To do this, this procedure should be followed:

1. → **Transform** → **Compute ...** [opens the **Compute Variable** dialog box shown in Plate 16.8]
2. under **Target Variable**: type **totalmin**

3. from the list of variables at the left, → **cardmins** [puts **cardmins** in box beneath **Numeric Expression**:]
→ **+** → **weimins** [puts **weimins** after + sign] → **+** button; → **othmins** [puts **othmins** after + sign]

4. → **OK**

The new variable **totalmin** will be created and will appear in the **Data Editor**.

Now at last, we can begin to analyse the data!



Data analysis with SPSS

Generating a frequency table

To produce a frequency table like the one in Table 15.2:

1. → **Analyze** → **Descriptive Statistics** → **Frequencies** ... [opens the **Frequencies** dialog box shown in Plate 16.9]

2. → **reasons for visiting** → **+** [puts **reasons for visiting** in **Variable[s]:** box]

3. → **OK**

The table will appear in the **Output Viewer** (see Plate 16.10).

Plate 16.9

The Frequencies dialog box

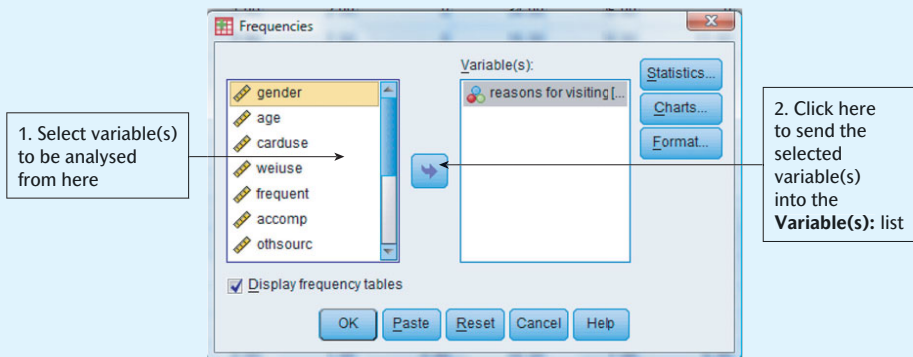
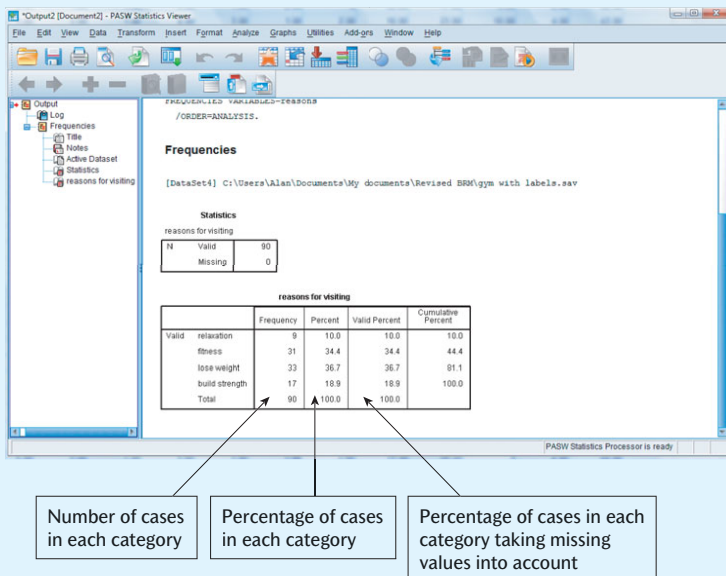


Plate 16.10

The Output Viewer with Frequency table



Note that in the **Frequencies** dialog box, variables that have been assigned labels will appear in terms of their variable labels, but those that have not been assigned labels will appear in terms of their variable names. This is a feature of all dialog boxes produced via **Analyze** and **Graphs** (see below).

Generating a bar chart

To produce a bar chart like the one in Figure 15.2:

1. → **Graphs** → **Chart Builder** . . . [opens **Chart Builder** dialog box shown in Plate 16.11]
2. → **Bar** below **Choose from:** and then → the simple bar format in the top left-hand corner of the **Gallery** and drag and drop it into the area above it. Then → **age** and drag and drop in the same way as for a bar chart.
3. → **reasons for visiting** from below **Variables:** and drag and drop into area marked in blue **X-Axis**?
4. → **OK**

Generating a pie chart

To produce a pie chart like the one in Figure 15.3:

1. → **Graphs** → **Chart Builder** . . . [opens the **Chart Builder** dialog box shown in Plate 16.12] → **Pie/Polar** below **Choose from:** and then → the pie chart format in the top left-hand corner of the **Gallery** and drag and drop it into the area above it.
2. → **reasons for visiting** from below **Variables:** and drag and drop into area marked in blue **Slice by?**
3. → **OK**

In order to include percentages, as in Figure 15.3, *double-click* anywhere in the chart in order to bring up the **Chart Editor**. The chart will appear in the **Chart Editor** and the main figure will become opaque. Then → **Elements** and then → **Show Data Labels**. The **Properties** sub-dialog box will appear (see Plate 16.12). Then to have labels and percentages as in Figure 15.3, rather

than frequencies ('counts'), which is the default, select **Percent**.

Your chart will be in colour, but, if you have access only to a monochrome printer, you can change your pie chart into patterns, which allows the slices to be clearer. This can be done through the **Chart Editor**.

Generating a histogram

In order to generate a histogram for an interval/ratio variable like **age**, → **Graphs** → **Histogram** below **Choose from:** and then → the histogram format in the top left-hand corner of the **Gallery** and drag and drop it into the area above it. Then → **age** and drag and drop it in the same way as for a bar chart. This procedure will generate a histogram whose age bands are defined by the software. By double-clicking on the diagram, the histogram can be edited using the **Chart Editor**. For example, colours can be changed or patterns inserted.

Generating the arithmetic mean, median, standard deviation, the range, and boxplots

To produce the mean, median, standard deviation, and the range for an interval/ratio variable like **age**, the following steps should be followed:

1. → **Analyze** → **Descriptive Statistics** → **Explore** . . . [opens the **Explore** dialog box]
2. → **age** →  to the left of **Dependent List:** [puts **age** in the **Dependent List:** box] → **Statistics** under **Display** → **OK**

The output will also include the 95 per cent confidence interval for the mean, which is based on the standard error of the mean. The output can be found in Table 16.1. If you select **Plots** . . . , the **Explore: Plots** sub-dialog box will come up and you can elect to generate a histogram. To do this, you will need to select either **Both** or **Plots** under **Display** on the **Explore** dialog box. In addition, selecting **Both** or **Plots** will produce two further types of figure, one of which is a boxplot, which was covered in Chapter 15.

Plate 16.11

Creating a bar chart with the Chart Builder

3. Select variable from here and drag and drop here

1. Select type of chart from here (in this case **Bar**)

2. Select format of graph here and drag and drop to here

Chart Builder

Variables:

- gender
- age
- reasons for visiti...
- carduse
- weiluse
- frequent
- accomp
- othsourc
- exercise

relaxation

fitness

Chart preview uses example data

Count

relaxation fitness [More...]

reasons for visiting

Gallery Basic Elements Groups/Point ID Titles/Footnotes

Choose from:

Favorites

- Bar
- Line
- Area
- Pie/Polar
- Scatter/Dot
- Histogram
- High-Low
- Boxplot
- Dual Axes

OK Paste Reset Cancel Help

Element Properties...

Options...

Plate 16.12

Creating a pie chart with the Chart Builder and Properties box

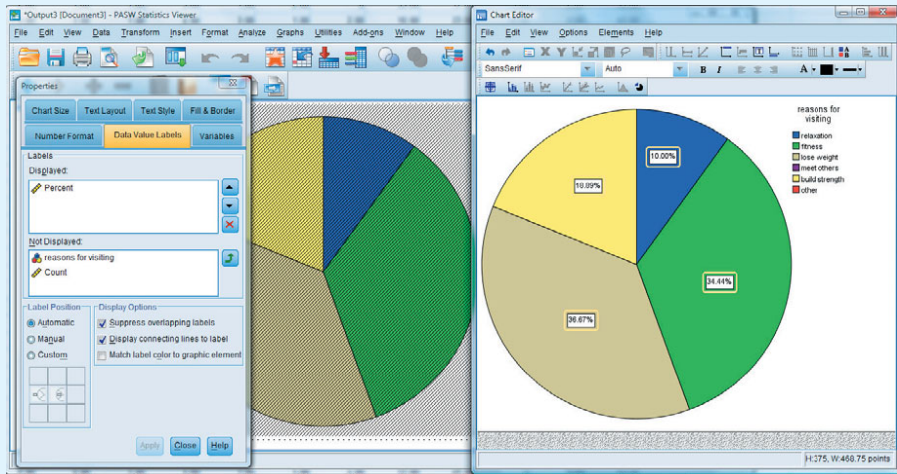


Table 16.1

Explore output for age (SPSS output)

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
age	89	98.9%	1	1.1%	90	100.0%

Descriptives

			Statistic	Std. Error
age	Mean		33.5955	.94197
	95% Confidence Interval for Mean	Lower Bound	31.7235	
		Upper Bound	35.4675	
	5% Trimmed Mean		33.3159	
	Median		31.0000	
	Variance		78.971	
	Std. Deviation		8.88656	
	Minimum		18.00	
	Maximum		57.00	
	Range		39.00	
	Interquartile Range		14.00	
	Skewness		.446	.255
	Kurtosis		-.645	.506

Generating a contingency table, chi-square, and Cramér's V

In order to generate a contingency table, like that in Table 15.4, along with a chi-square test and Cramér's V, the following procedure should be followed:

1. → **Analyze** → **Descriptive Statistics** → **Crosstabs...** [opens the Crosstabs dialog box shown in Plate 16.13]
2. → **reasons for visiting** → **by Row[s]** [reasons for visiting will appear in the Row[s]: box] →

Plate 16.13

The Crosstabs dialog box

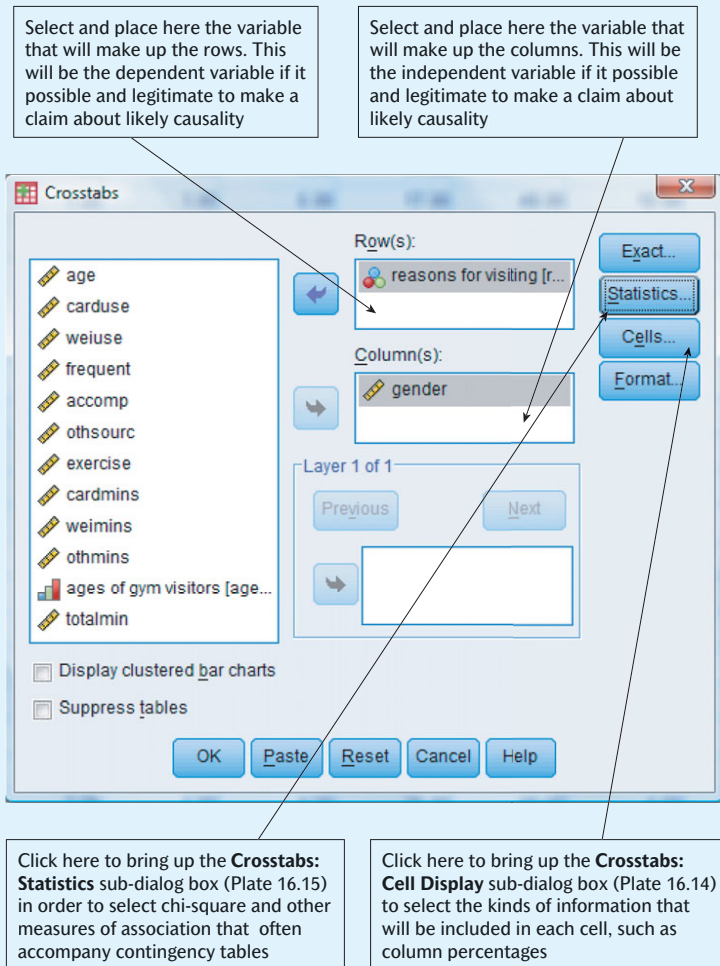
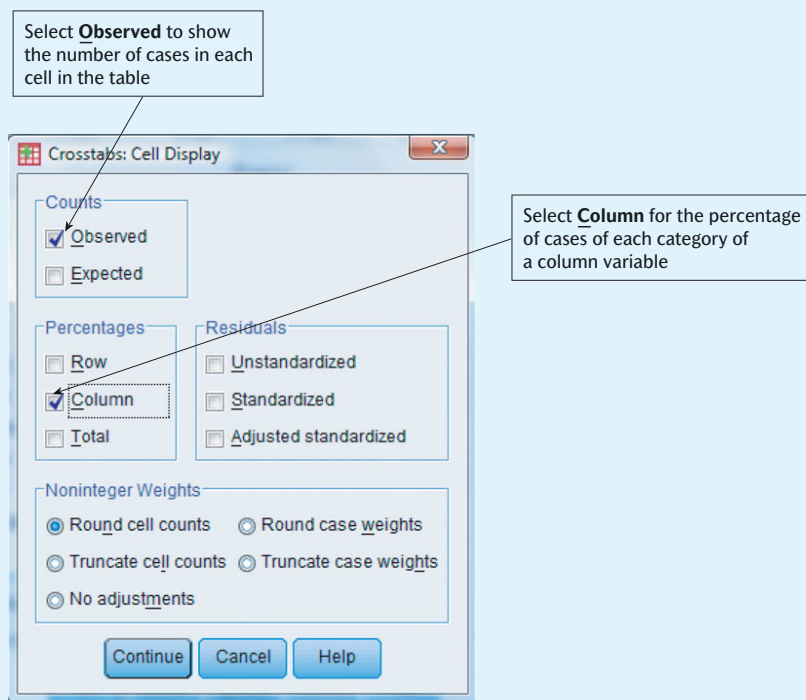


Plate 16.14

The Crosstabs: Cell Display sub-dialog box



gender → by Column[s]: [*gender* will appear in the Column[s]: box] → Cells... [opens **Crosstabs: Cell Display** sub-dialog box shown in Plate 16.14]

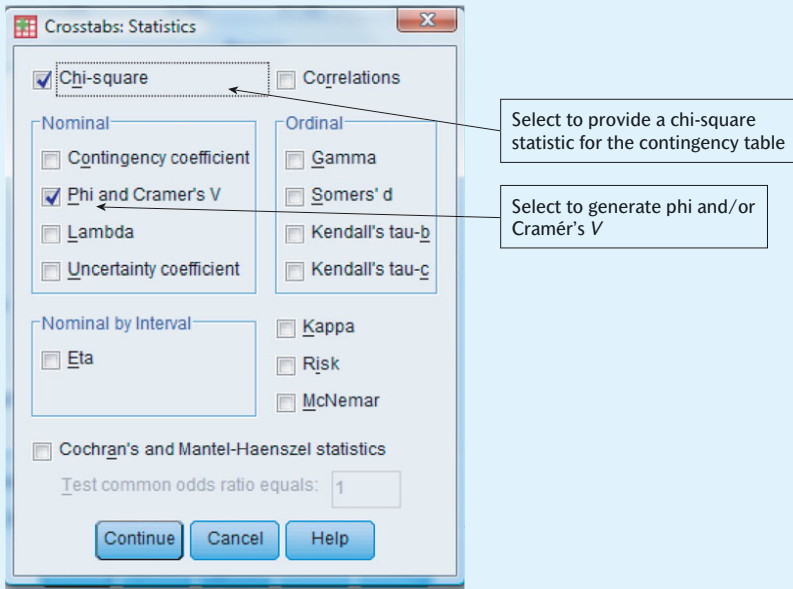
3. Make sure **O**bserved in the **C**ounts box has been selected. Make sure Column under **P**ercentages has been selected. If either of these has not been selected, simply click at the relevant point. → **C**ontinue [closes **Crosstabs: Cell Display** sub-dialog box and returns you to the **Crosstabs** dialog box shown in Plate 16.13]
4. → **S**tatistics... [opens the **Crosstabs: Statistics** sub-dialog box shown in Plate 16.15]
5. → **C**hi-square → **P**hi and Cramér's *V* → **C**ontinue [closes **Crosstabs: Statistics** sub-dialog box and returns you to the **Crosstabs** dialog box shown in Plate 16.13]
6. → **O**K

The resulting output can be found in Table 16.2.

If you have a table with two dichotomous variables, you would use the same sequence of steps to produce phi.

Plate 16.15

The Crosstabs: Statistics sub-dialog box



Generating Pearson's r and Spearman's rho

To produce Pearson's r in order to find the correlations between *age*, *cardmins*, and *weimins*, follow these steps:

1. → **Analyze** → **Correlate** → **Bivariate...** [opens **Bivariate Correlations** dialog box shown in Plate 16.16]
2. → *age* → → *cardmins* → → *weimins* → [age, cardmins, and weimins should now be in the **Variables:** box] → **Pearson** [if not already selected] → **OK**

The resulting output is in Table 16.3.

To produce correlations with Spearman's rho, follow the same procedure, but, instead of selecting **Pearson**, you should → **Spearman** instead.

Generating scatter diagrams

Scatter diagrams, known as *scatterplots* in SPSS, are produced in the following way. Let us say that we want to plot the relationship between *age* and *cardmins*. There is a convention that, if one variable can be identified as likely to be the independent variable, it should be placed on the x axis—that is, the horizontal axis. Since *age* is bound to be the independent variable, we would follow these steps:

1. → **Graphs** → **Chart Builder** [opens the **Chart Builder** dialog box shown in Plate 16.17]
2. → **Scatter/Dot** from below **Choose from:**. Then select from the scatter diagram formats, the basic format which is in the top left-hand corner and drag and drop into the area above the scatter diagram formats
3. → *cardmins* and drag and drop into area designated **Y-Axis?** and → *age* and drag and drop into area designated **X-Axis?** (see Plate 16.17)

Table 16.2

Contingency table for *reasons for visiting by gender* (SPSS output)

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
reasons for visiting * gender	90	100.0%	0	.0%	90	100.0%

reasons for visiting * gender Crosstabulation

			gender		Total
			Male	Female	
reasons for visiting	relaxation	Count % within gender	3 7.1%	6 12.5%	9 10.0%
	fitness	Count % within gender	15 35.7%	16 33.3%	31 34.4%
	lose weight	Count % within gender	8 19.0%	25 52.1%	33 36.7%
	build strength	Count % within gender	16 38.1%	1 2.1%	17 18.9%
Total		Count % within gender	42 100.0%	48 100.0%	90 100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	22.726 ^a	3	.000
Likelihood Ratio	25.805	3	.000
Linear-by-Linear Association	9.716	1	.002
N of Valid Cases	90		

a. 2 cells (25.0%) have expected count less than 5.
The minimum expected count is 4.20.

Interpret the Pearson Chi-Square values for information about chi-square

Symmetric Measures

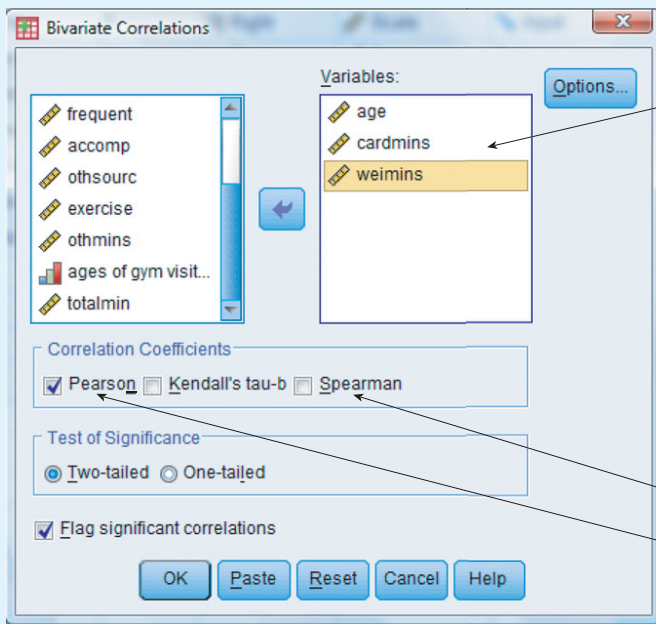
		Value	Approx. Sig.
Nominal by Nominal	Phi	.503	.000
	Cramer's V	.503	.000
N of Valid Cases		90	

Shows the strength of the relationship between the variables

Shows the level of statistical significance of the computed value of Cramér's V

Plate 16.16

The Bivariate Correlations dialog box



All variables to be correlated go here

Select **Pearson** or **Spearman** depending on the kinds of variables being analysed

Table 16.3

Correlations output for age, weimins, and cardmins (SPSS output)

Correlations

Correlations		age	cardmins	weimins
age	Pearson Correlation	1	-.109	-.273**
	Sig. (2-tailed)		.311	.010
	N	89	89	89
cardmins	Pearson Correlation	-.109	1	-.161
	Sig. (2-tailed)	.311		.130
	N	89	90	90
weimins	Pearson Correlation	-.273**	-.161	1
	Sig. (2-tailed)	.010	.130	
	N	89	90	90

Correlations of $p < 0.05$ are 'flagged' with asterisks

Shows number of cases involved in the calculation of a correlation, less any cases for which there are missing values for either or both variables

** . Correlation is significant at the 0.01 level (2-tailed)

Shows strength of relationship between variables as indicated by Pearson's r

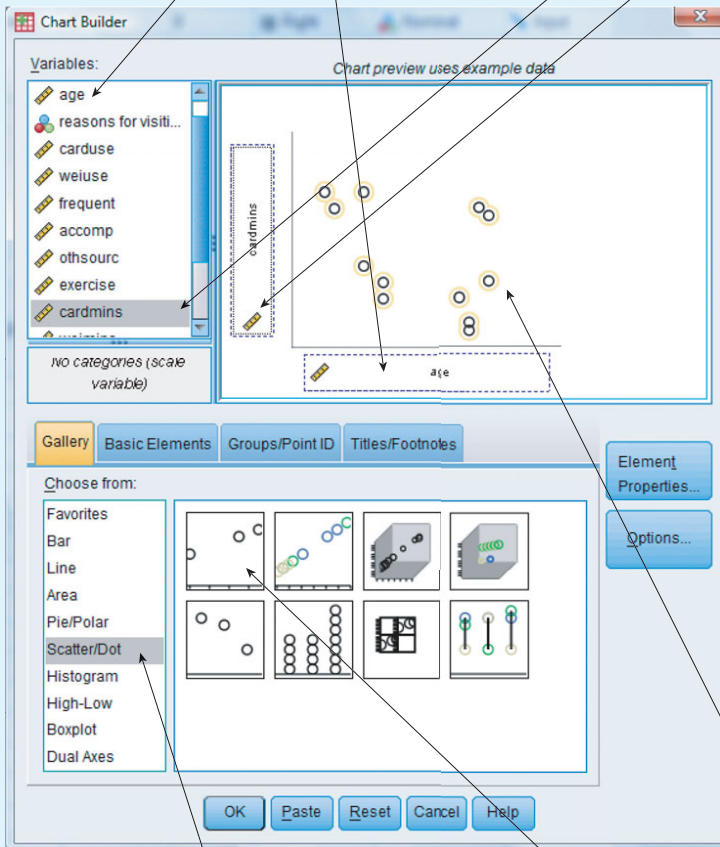
Shows level of statistical significance of computed value of Pearson's r

Plate 16.17

Creating a scatter diagram with the Chart Builder

3. Select independent variable from here and drag and drop to here

4. Select dependent variable from here and drag and drop to here

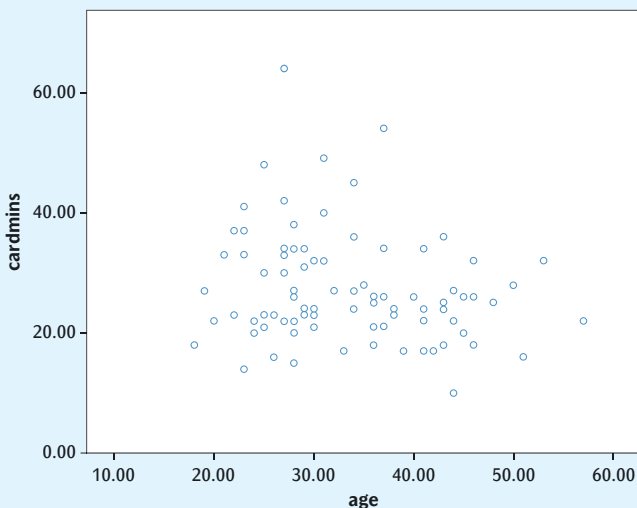


1. Select type of chart from here

2. Select format of scatterplot from here and drag and drop to here

Figure 16.1

Scatter diagram showing the relationship between *age* and *cardmins* (SPSS output)



A default scatter diagram is shown in Figure 16.1. The scatter diagram can then be edited by bringing up the **Chart Editor**. For example, the type and size of the markers can be changed by clicking anywhere in the chart in the **Chart Editor**. This brings up a **Properties** sub-dialog box, which allows a variety of changes to the appearance of the diagram, such as colour and the nature of the points on the plot.

Comparing means and eta

To produce a table like Table 15.5, these steps should be followed:

1. → **Analyze** → **Compare Means** → **Means** . . . [opens the **Means** dialog box shown in Plate 16.18]
2. → *cardmins* →  to the left of **Dependent List**: → reasons for visiting →  to the left of **Independent List**: → **Options** . . . [opens the **Means: Options** sub-dialog box]
3. → **Anova table and eta** underneath **Statistics for First Layer** → **Continue** [closes the **Means: Options**

sub-dialog box and returns you to the **Means** dialog box shown in Plate 16.18] → **OK**

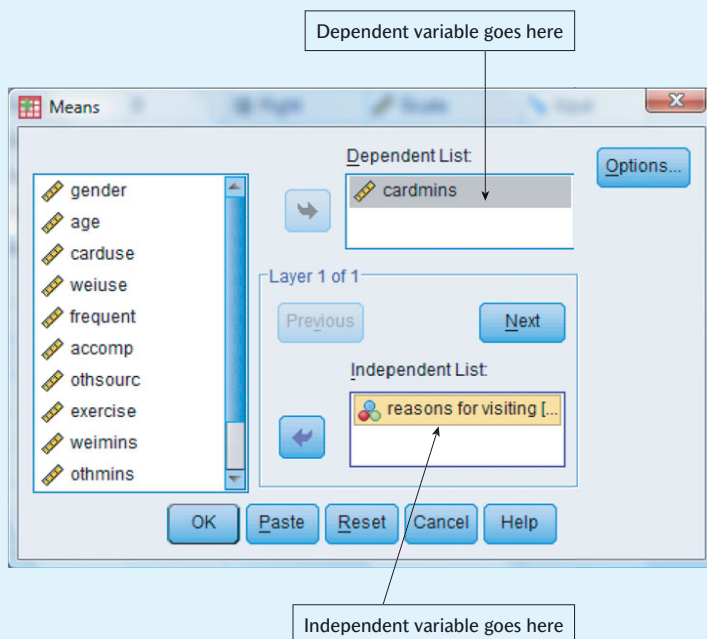
Generating a contingency table with three variables

To create a table like that in Table 15.7, you would need to follow these steps:

1. → **Analyze** → **Descriptive Statistics** → **Crosstabs** . . . [opens the **Crosstabs** dialog box shown in Plate 16.13]
2. → *othsourc* →  by **Row[s]** [*othsourc* will appear in the **Row[s]**: box]
3. → *age3* [this is the name we gave when we created a new variable with *age* recoded into three categories] →  by **Column[s]**: [*age3* will appear in the **Column[s]**: box] → *gender* →  beneath **Previous** [*gender* will appear in the box underneath **Layer 1 of 1**] → **Cells** [opens **Crosstabs: Cell Display** sub-dialog box shown in Plate 16.14]
4. Make sure **Observed** in the **Counts** box has been selected. Make sure **Column** under **Percentages** has

Plate 16.18

The Means dialog box



been selected. If either of these has not been selected, simply click at the relevant point. → **Continue** [closes **Crosstabs: Cell Display** sub-dialog box and returns you to the **Crosstabs** dialog box shown in Plate 16.13]

5. → **OK**

The resulting table will look somewhat different from Table 15.7 in that gender will appear as a row rather than as a column variable.



Further operations in SPSS

Saving your data

You will need to save your data for future use. To do this, make sure that the **Data Editor** is the active window. Then,

→ **File** → **Save As...**

The **Save Data As** dialog box will then appear. You will need to provide a name for your data, which will be placed after **File name:** We called the file 'gym study'. You also need to decide where you are going to save the data—for example, onto a memory stick. To select the destination drive, → the downward pointing arrow to

the left of **Look in** and then select the drive and folder into which you want to place your data. Then → **Save**.

Remember that this procedure saves your data *and* any other work you have done on your data—for example, value labels and recoded variables. If you subsequently use the data again and do more work on your data, such as creating a new variable, you will need to save the data again or the new work will be lost. SPSS will give you a choice of renaming your data, in which case you will have two files of data (one with the original data and one with any changes), or keeping the same name, in which case the file will be changed and the existing name retained.

Retrieving your data

When you want to retrieve the data file you have created, → **File** → **Open . . .** The **Open File** dialog box will appear. You then need to go to the location in which you have deposited your data to retrieve the file containing

your data and then → **Open** → **Data . . .** A shortcut alternative to this procedure is to → the first button on the toolbar (it looks like an open file), which brings up the **Open File** dialog box.

Printing output

To print all the output in the **SPSS Output Viewer**, make sure that the **Output 1 – SPSS Viewer** is the active window and then → **File** → **Print . . .** The **Print** dialog box will appear and then → **OK**. To print just some of your output, hold down the Ctrl button on your keyboard and click once on the parts you want to print. The easiest way to do this is to select all the elements you want in the output summary in the left-hand segment of the **Output Viewer** shown in Plate 16.10. Then bring up the **Print** dialog box. When the **Print** dialog box appears, make sure **Selection** under **Print range** has been selected. The third button on the toolbar (which appears as a printer) provides a shortcut to the **Print** dialog box.



Key points

- SPSS can be used to implement the techniques learned in Chapter 15, but learning new software requires perseverance and at times the results obtained may not seem to be worth the learning process.
- But it is worth it—it would take you far longer to perform calculations on a sample of around 100 than to learn the software.
- If you find yourself moving into much more advanced techniques, the time saved is even more substantial, particularly with large samples.
- It is better to become familiar with SPSS before you begin designing your research instruments, so you are aware of difficulties you might have in presenting your data in SPSS at an early stage.



Questions for review

Getting started in SPSS

- Outline the differences between: variable names, variable labels, and value labels.
- In what circumstances might you want to recode a variable?
- In what circumstances might you want to create a new variable?

Data analysis with SPSS

Using the gym survey data, create:

- a frequency table for **exercise**;
- a bar chart and pie chart for **exercise** and compare their usefulness;

- a histogram for **cardmins**;
- measures of central tendency and dispersion for **cardmins**;
- a contingency table and chi-square test for **exercise** and **gender**;
- Pearson's r for **age** and **cardmins**;
- Spearman's rho for **carduse** and **weiuise**;
- a scatter diagram for **age** and **cardmins**;
- a comparing means analysis for **totalmin** and **reasons for visiting**.



Online Resource Centre

www.oxfordtextbooks.co.uk/orc/brymansrm4e/

Visit the Online Resource Centre that accompanies this book to enrich your understanding of using SPSS for Windows. Consult web links, test yourself using multiple choice questions, and gain further guidance and inspiration from the Student Researcher's Toolkit.
